Unsupervised Degradation Representation Learning for Blind Super-Resolution

Longguang Wang[1], Yingqian Wang[1], Xiaoyu Dong[2,3], Qingyu Xu[1], Jungang Yang[1], Wei An[1], Yulan Guo[1*]

[1]National University of Defense Technology   [2]The University of Tokyo   [3]RIKEN AIP

{wanglongguang15,yulan.guo}@nudt.edu.cn

Abstract

Most existing CNN-based super-resolution (SR) methods are developed based on an assumption that the degradation is fixed and known (e.g., bicubic downsampling). However, these methods suffer a severe performance drop when the real degradation is different from their assumption. To handle various unknown degradations in real-world applications, previous methods rely on degradation estimation to reconstruct the SR image. Nevertheless, degradation estimation methods are usually time-consuming and may lead to SR failure due to large estimation errors. In this paper, we propose an unsupervised degradation representation learning scheme for blind SR without explicit degradation estimation. Specifically, we learn abstract representations to distinguish various degradations in the representation space rather than explicit estimation in the pixel space. Moreover, we introduce a Degradation-Aware SR (DASR) network with flexible adaption to various degradations based on the learned representations. It is demonstrated that our degradation representation learning scheme can extract discriminative representations to obtain accurate degradation information. Experiments on both synthetic and real images show that our network achieves state-of-the-art performance for the blind SR task. Code is available at: https://github.com/LongguangWang/DASR.

1. Introduction

Single image super-resolution (SR) aims at recovering a high-resolution (HR) image from a low-resolution (LR) observation. Recently, CNN-based methods [9, 22, 24, 2, 32] have dominated the research of SR due to the powerful feature representation capability of deep neural networks. As a typical inverse problem, SR is highly coupled with the degradation model [3]. Most existing CNN-based methods are developed based on an assumption that the degradation is known and fixed (*e.g.*, bicubic downsampling). However, these methods suffer a severe performance drop when the real degradation differs from their assumption [12].
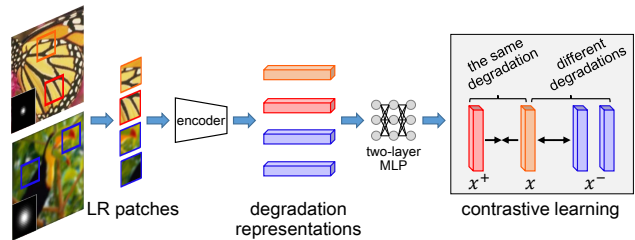

Figure 1. An illustration of our unsupervised degradation representation learning scheme.

To handle various degradations in real-world applications, several methods [42, 33, 40, 34] have been proposed to investigate the non-blind SR problem. Specifically, these methods use a set of degradations (*e.g.*, different combinations of Gaussian blurs, motion blurs and noises) for training and assume the degradation of the test LR image is known at the inference time. These non-blind methods produce promising SR results when the true degradation is known in priori.

To super-resolve real images with unknown degradations, degradation estimation [28, 3] needs to be performed to provide degradation information for non-blind SR networks [42, 33, 40, 34]. However, these non-blind methods are sensitive to degradation estimation. Consequently, the estimation error can further be magnified by the SR network, resulting in obvious artifacts [12]. To address this problem, Gu *et al.* [12] proposed an iterative kernel correction (IKC) method to correct the estimated degradation by observing previous SR results. By iteratively correcting the degradation, artifact-free results can be gradually produced. Since numerous iterations are required at test time by degradation estimation methods [28, 3] and IKC [12], these methods are time-consuming.

Unlike the above methods that explicitly estimate the degradation from an LR image, we investigate a different approach by learning a degradation representation to distinguish the latent degradation from other ones. Motivated by recent advances of contrastive learning [13, 10, 35, 17, 5], a contrastive loss is used to conduct unsupervised degradation representation learning by contrasting positive pairs against negative pairs in the latent space (Fig. 1). The

benefits of degradation representation learning are twofold: **First**, compared to extracting full representations to estimate degradations, it is easier to learn abstract representations to distinguish different degradations. Consequently, we can obtain a discriminative degradation representation to provide accurate degradation information at a single inference. **Second**, degradation representation learning does not require the supervision from groundtruth degradation. Thus, it can be conducted in an unsupervised manner and is more suitable for real-world applications with unknown degradations.

In this paper, we introduce an unsupervised degradation representation learning scheme for blind SR. Specifically, we assume the degradation is the same in an image but can vary for different images, which is the general case widely used in literature [42, 3, 40]. Consequently, an image patch should be similar to other patches in the same image (*i.e.*, with the same degradation) and dissimilar to patches from other images (*i.e.*, with different degradations) in the degradation representation space, as illustrated in Fig. 1. Moreover, we propose a degradation-aware SR (DASR) network with flexible adaption to different degradations based on the learned representations. Specifically, our DASR incorporates degradation information to perform feature adaption by predicting convolutional kernels and channel-wise modulation coefficients from the degradation representation. Experimental results show that our network can handle various degradations and produce promising results on both synthetic and real-world images under blind settings.

2. Related Work

In this section, we briefly review several major works for CNN-based single image SR and recent advances of contrastive learning.

2.1. Single Image Super-Resolution

SR with Single Degradation. As a pioneer work, a three-layer network is used in SRCNN [9] to learn the LR-HR mapping for single image SR. Since then, CNN-based methods have dominated the research of SR due to their promising performance. Kim *et al.* [21] proposed a 20-layer network with a residual learning strategy. Lim *et al.* [24] followed the idea of residual learning and modified residual blocks to build a very deep and wide network, namely EDSR. Zhang *et al.* [45] then combined residual learning and dense connection to construct a residual dense network (RDN) with over 100 layers. Haris *et al.* [14] introduced multiple up-sampling/down-sampling layers to provide an error feedback mechanism and used self-corrected features to produce superior results. Recently, channel attention and second-order channel attention are further introduced by RCAN [44] and SAN [7] to exploit feature correlation for improved performance.

SR with Multiple Degradations. Although important advances have been achieved by the above SR methods, they are tailored to a fixed bicubic degradation and suffer severe performance drop when the real degradation differs from the bicubic one [12]. To handle various degradations, several efforts [42, 38, 40, 20] have been made to investigate the non-blind SR problem. Specifically, degradation is first used as an additional input in SRMD [42] to super-resolve LR images under different degradations. Later, dynamic convolutions are further incorporated in UDVD [38] to achieve better performance than SRMD. Recently, Zhang *et al.* [40] developed an unfolding SR network (USRnet) to handle different degradations by alternately solving a data sub-problem and a prior sub-problem. Hussein *et al.* [20] introduced a closed-form correction filter to transform an LR image to match the one generated by bicubic degradation. Then, existing networks trained for bicubic degradation can be used to super-resolve the transformed LR image.

Zero-shot methods have also been investigated to achieve SR with multiple degradations. In ZSSR [33], training is conducted at test time using a degradation and an LR image as its input. Consequently, the network can be adapted to the given degradation. However, ZSSR requires thousands of iterations to converge and is quite time-consuming. To address this limitation, optimization-based meta-learning is used in MZSR [34] to make the network adaptive to a specific degradation within a few iterations during inference.

Since degradation is used as an input for these aforementioned methods, they highly rely on degradation estimation methods [28, 3] for blind SR. Therefore, degradation estimation errors can ultimately introduce undesired artifacts to the SR results [12]. To address this problem, Gu *et al.* [12] proposed an iterative kernel correction (IKC) method to correct the estimated degradation by observing previous SR results. Luo *et al.* [25] developed a deep alternating network (DAN) by iteratively estimating the degradation and restoring an SR image.

2.2. Contrastive Learning

Contrastive learning has demonstrated its effectiveness in unsupervised representation learning. Previous methods [8, 43, 29, 11] usually conduct representation learning by minimizing the difference between the output and a fixed target (*e.g.*, the input itself for auto-encoders). Instead of using a pre-defined and fixed target, contrastive learning maximizes the mutual information in a representation space. Specifically, the representation of a query sample should attract positive counterparts while repelling negative counterparts. The positive counterparts can be transformed versions of the input [37, 5, 17], multiple views of the input [35] and neighboring patches in the same image [30, 18]. In this paper, image patches generated with the same degrada-

tion are considered as positive counterparts and contrastive learning is conducted to obtain content-invariant degradation representations, as shown in Fig. 1.

3. Methodology

3.1. Problem Formulation

The degradation model of an LR image I^{LR} can be formulated as follows:

$$I^{LR} = (I^{HR} \otimes k) \downarrow_s + n, \qquad (1)$$

where I^{HR} is the HR image, k is a blur kernel, $\otimes$ denotes convolution operation, $\downarrow_s$ represents downsampling operation with scale factor s and n usually refers to additive white Gaussian noise. Following [42, 12], we use bicubic downsampler as the downsampling operation. In this paper, we first investigate a noise-free degradation model with isotropic Gaussian kernels and then a more general degradation model with anisotropic Gaussian kernels and noises. Finally, we test our network on real-world degradations.

3.2. Our Method

Our blind SR framework consists of a degradation encoder and a degradation-aware SR network, as illustrated in Fig. 2. First, the LR image is fed to the degradation encoder (Fig. 2(a)) to obtain a degradation representation. Then, this representation is incorporated in the degradation-aware SR network (Fig. 2(b)) to produce the SR result.

3.2.1 Degradation Representation Learning

The goal of degradation representation learning is to extract a discriminative representation from the LR image in an unsupervised manner. As shown in Fig. 1, we use a contrastive learning framework [17] for degradation representation learning. Note that, we assume the degradation is the same in each image and varies for different images.

Formulation. Given an image patch (annotated with an orange box in Fig. 1) as the query patch, other patches extracted from the same LR image (*e.g.*, the patch annotated with a red box) can be considered as positive samples. In contrast, patches from other LR images (*e.g.*, patches annotated with blue boxes) can be referred to as negative samples. Then, we encode the query, positive and negative patches into degradation representations using a convolutional network with six layers (Fig. 2(a)). As suggested in SimCLR [5] and MoCo v2 [6], the resulting representations are further fed to a two-layer multi-layer perceptron (MLP) projection head to obtain x, x^+ and x^-. x is encouraged to be similar to x^+ while being dissimilar to x^-. Following MoCo [17], an InfoNCE loss is used to measure the simi-

larity. That is,

$$L_x = -\log \frac{\exp(x \cdot x^+/\tau)}{\sum_{n=1}^{N} \exp(x \cdot x_n^-/\tau)}, \qquad (2)$$

where N is the number of negative samples, τ is a temperature hyper-parameter and $\cdot$ represents the dot product between two vectors.

As emphasized in existing contrastive learning methods [5, 17, 31], a large dictionary covering a rich set of negative samples is critical for good representation learning. To obtain content-invariant degradation representations, a queue containing samples with various contents and degradations is maintained. During training, B LR images (*i.e.*, B different degradations) are first randomly selected and then two patches are randomly cropped from each image. Next, these $2B$ patches are encoded into $\{p_i^1, p_i^2 \in \mathbb{R}^{256}\}$ using our degradation encoder, where p_i^1 is the embedding of the first patch from the ith image. For the ith image, we refer to p_i^1 and p_i^2 as query and positive samples. The overall loss is defined as:

$$L_{degrad} = \sum_{i=1}^{B} -\log \frac{\exp(p_i^1 \cdot p_i^2/\tau)}{\sum_{j=1}^{N_{queue}} \exp(p_i^1 \cdot p_{queue}^j/\tau)}, \quad (3)$$

where N_{queue} is the number of samples in the queue and p_{queue}^j represents the jth negative sample.

Discussion. Existing degradation estimation methods [28, 3, 12] aim at estimating the degradation (usually the blur kernel) at pixel level. That is, these methods learn to extract full representations of the degradation. However, they are time-consuming as numerous iterations are required during inference. For example, KernelGAN conducts network training during test and takes over 60 seconds for a single image [3]. Different from these methods, we aim at learning a "good" abstract representation to distinguish a specific degradation from others rather than explicitly estimating the degradation. It is demonstrated in Sec. 4.2 that our degradation representation learning scheme is effective yet efficient and can obtain discriminative representations at a single inference. Moreover, our scheme does not require the supervision from groundtruth degradation and can be conducted in an unsupervised manner.

3.2.2 Degradation-Aware SR Network

With degradation representation learning, a degradation-aware SR (DASR) network is proposed to super-resolve the LR image using the resultant representation, as shown in Fig. 2(b).

Network Architecture. Figure 2(b) illustrates the architecture of our DASR network. Degradation-aware block (DA block) is used as the building block and the high-level structure of RCAN [44] is employed. Our DASR network con-
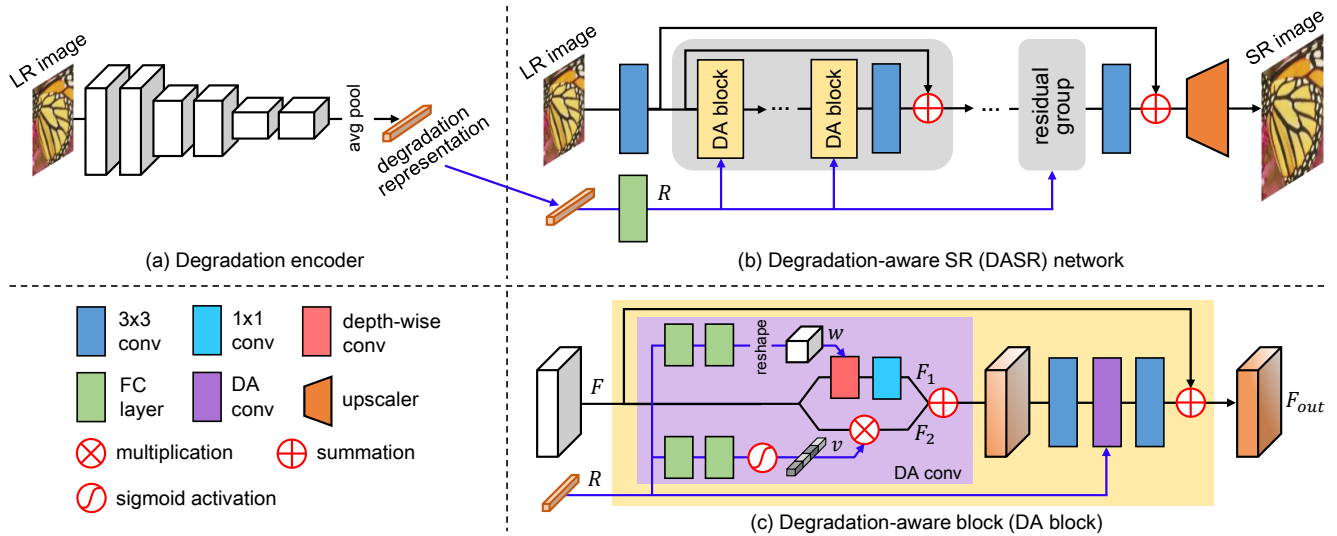
Figure 2. An overview of our blind SR framework.

sists of 5 residual groups, with each group comprising of 5 DA blocks.

Within each DA block, two DA convolutional layers are used to adapt the features based on the degradation representation, as shown in Fig. 2(c). Motivated by the observation that convolutional kernels of models trained for different restoration levels share similar patterns but have different statistics [15], our DA convolutional layer learns to predict the kernel of a depth-wise convolution conditioned on the degradation representation. Specifically, the degradation representation R is fed to two full-connected (FC) layers and a reshape layer to produce a convolutional kernel $w \in \mathbb{R}^{C \times 1 \times 3 \times 3}$. Then, the input feature F is processed with a 3×3 depth-wise convolution (using w) and a 1×1 convolution to produce F_1. Moreover, inspired by CResMD [16] (that uses controlling variables to rescale different channels to handle multiple degradations), our DA convolutional layer also learns to generate modulation coefficients based on the degradation representation to perform channel-wise feature adaption. Specifically, R is passed to another two FC layers and a sigmoid activation layer to generate channel-wise modulation coefficients v. Then, v is used to rescale different channel components in F, resulting in F_2. Finally, F_1 is summed up with F_2 and fed to the subsequent layers to produce the output feature F_{out}.

Discussion. Existing SR networks [42, 38] for multiple degradations commonly concatenate degradation representations with image features and feed them to CNNs to exploit degradation information. However, due to the domain gap between degradation representations and image features, directly processing them as a whole using convolution will introduce interference [12]. Different from these networks, by learning to predict convolutional kernels and modulation coefficients based on the degradation represen-

tations, our DASR can well exploit degradation information to adapt to specific degradations. It is demonstrated in Sec. 4.2 that our DASR benefits from DA convolution to achieve flexible adaption to various degradations with better SR performance.

4. Experiments

4.1. Datasets and Implementation Details

We synthesized LR images according to Eq. 1 for training and test. Following [12], we used 800 training images in DIV2K [1] and 2650 training images in Flickr2K [36] as the training set, and included four benchmark datasets (Set5 [4], Set14 [39], B100 [27] and Urban100 [19]) for evaluation. The size of the Gaussian kernel was fixed to 21×21 following [12]. We first trained our network on noise-free degradations with isotropic Gaussian kernels only. The ranges of kernel width σ were set to [0.2,2.0], [0.2,3.0] and [0.2,4.0] for $\times 2/3/4$ SR, respectively. Then, our network was trained on more general degradations with anisotropic Gaussian kernels and noises. Anisotropic Gaussian kernels characterized by a Gaussian probability density function $N(0, \Sigma)$ (with zero mean and varying covariance matrix Σ) were considered. The covariance matrix Σ was determined by two random eigenvalues $\lambda_1, \lambda_2 \sim U(0.2, 4)$ and a random rotation angle $\theta \sim U(0, \pi)$. The range of noise level was set to $[0, 25]$.

During training, 32 HR images were randomly selected and data augmentation was performed through random rotation and flipping. Then, we randomly chose 32 Gaussian kernels from the above ranges to generate LR images. For general degradations, Gaussian noises were also added to the resultant LR images. Next, 64 LR patches of size 48×48 (two patches from each LR image as illustrated in

Table 1. PSNR results achieved on Set14 for ×4 SR. *: SRMDNF is a version of SRMD trained with noise-free samples.

Method	Degradation Representation Learning	DA Conv		Oracle Degradation	Time	Kernel Width σ				
		Kernel	Modulation			0.2	1.0	1.8	2.6	3.4
SRMDNF* [42]	-	-	-	✓	3ms	28.44	28.50	28.49	28.31	27.55
SRMDNF [42] + KernelGAN [3]	-	-	-	✗	3ms+190s	26.62	26.74	26.62	26.88	26.66
SRMDNF [42] + Predictor [12]	-	-	-	✗	3ms+2ms	26.13	26.15	26.19	26.20	26.18
Model 1	✗	✓	✓	✗	70ms	28.46	28.40	28.30	27.77	26.79
Model 2	✓	✗	✗	✗	51ms	28.49	28.38	27.99	27.54	26.72
Model 3	✓	✓	✗	✗	67ms	28.42	28.30	28.21	27.97	27.33
Model 4 (Ours)	✓	✓	✓	✗	70ms	28.50	28.45	28.40	28.16	27.58
Model 5	-	✓	✓	✓	61ms	28.60	28.67	28.69	28.48	27.90

Sec. 3.2.1) and their corresponding HR patches were randomly cropped. In our experiments, we set τ and N_{queue} in Eq. 3 to 0.07 and 8192, respectively. The Adam method [23] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ was used for optimization. We first trained the degradation encoder by optimizing L_{degrad} for 100 epochs. The initial learning rate was set to 1×10^{-3} and decreased to 1×10^{-4} after 60 epochs. Then, we trained the whole network for 500 epochs. The initial learning rate was set 1×10^{-4} and decreased to half after every 125 epochs. The overall loss function is defined as $L = L_{SR} + L_{degrad}$, where L_{SR} is the L_1 loss between SR results and HR images.

4.2. Experiments on Noise-Free Degradations with Isotropic Gaussian Kernels

We first conduct ablation experiments on noise-free degradations with only isotropic Gaussian kernels. Then, we compare our DASR to several recent SR networks, including RCAN [44], SRMD [42], MZSR [33] and IKC [12]. RCAN is a state-of-the-art PSNR-oriented SR method for bicubic degradation. MZSR is a non-blind zero-shot SR method for degradations with isotorpic/anisotropic Gaussian kernels. SRMD is a non-blind SR method for degradations with isotropic/anisotropic Gaussian kernels and noises. IKC is a blind SR method that only considers degradations with isotropic Gaussian kernels. Note that, we do not include DAN [25], USRnet [40] and correction filter [20] for comparison since their degradation model is different from ours. These methods use s-fold downsampler[1] rather than bicubic downsampler as the downsampling operation in Eq. 1. To achieve fair comparison with [25, 40, 20], we re-trained our DASR using their degradation model and provide the results in the supplemental material.

Degradation Representation Learning. Degradation representation learning is used to produce discriminative representations to provide degradation information. To demonstrate its effectiveness, we introduced a network variant (Model 1) by removing degradation representation learning. Specifically, L_{degrad} was excluded during training without changing the network. Besides, the separate training of degradation encoder was removed and the whole network was directly trained for 500 epochs.

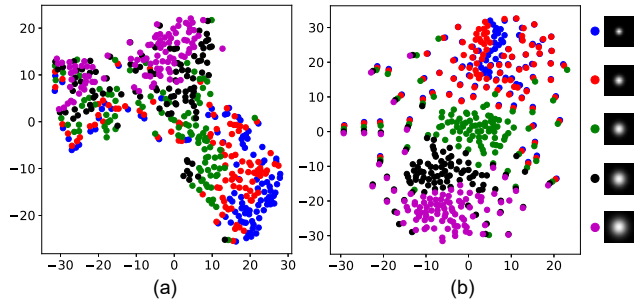
[1]Extract upper-left pixel within each $s \times s$ patch.

Figure 3. Visualization of representations for degradations with different kernel widths σ. (a) illustrates representations generated by our DASR w/o degradation representation learning (Model 1). (b) illustrates representations generated by our DASR (Model 4).

We first compare the degradation representations learned by models 1 and 4. Specifically, we used B100 to generate LR images with different degradations and fed them to models 1 and 4 to produce degradation representations. Then, these representations are visualized using the T-SNE method [26]. It can be observed in Fig. 3(b) that our degradation representation learning scheme can generate discriminative clusters. Without degradation representation learning, degradations with various kernel widths cannot be well distinguished, as shown in Fig. 3(a). This demonstrates that degradation representation learning facilitates our degradation encoder to learn discriminative representations to provide accurate degradation information. We further compare the SR performance of models 1 and 4 in Table 1. If degradation representation learning is removed, model 1 cannot handle multiple degradations well and produces lower PSNR values, especially for large kernel widths. In contrast, model 4 benefits from accurate degradation information provided by degradation representation learning to achieve better SR performance.

Degradation-Aware Convolutions. With degradation encoder, the extracted degradation representation is incorporated by DA convolutions to achieve flexible adaption to different degradations by predicting convolutional kernels and channel-wise modulation coefficients. To demonstrate the effectiveness of these two key components, we first introduced a variant (Model 2) by replacing DA convolutions with vanilla ones. Specifically, degradation representations are stretched and concatenated with image features as in [42] before being fed to vanilla convolutions. Then, we de-

Table 2. PSNR results achieved on noise-free degradations with isotropic Gaussian kernels. Note that, the degradation becomes bicubic degradation when kernel width is set to 0. Running time is averaged on Set14.

Method	Scale	Time	Set5				Set14				B100				Urban100			
Kernel Width			0	0.6	1.2	1.8	0	0.6	1.2	1.8	0	0.6	1.2	1.8	0	0.6	1.2	1.8
Bicubic		-	33.66	32.30	29.28	27.07	30.24	29.21	27.13	25.47	29.56	28.76	26.93	25.51	26.88	26.13	24.46	23.06
RCAN [44]		170ms	**38.27**	35.91	31.20	28.50	**34.12**	32.31	28.48	26.33	**32.41**	31.16	28.04	26.26	**33.34**	29.80	25.38	23.44
SRMDNF [42] + Predictor [12]	×2	9ms	34.94	34.77	34.13	33.80	31.48	31.35	30.78	30.18	30.77	30.33	29.89	29.20	29.05	28.42	27.43	27.12
MZSR [34] + Predictor [12]		90ms	35.96	35.66	35.22	32.32	31.97	31.33	30.85	29.17	30.64	29.82	29.41	28.72	29.49	29.01	28.43	26.39
DASR (Ours)		71ms	37.87	**37.47**	**37.19**	**35.43**	33.34	**32.96**	**32.78**	**31.60**	32.03	**31.78**	**31.71**	**30.54**	31.49	**30.71**	**30.36**	**28.95**
Kernel Width			0	0.8	1.6	2.4	0	0.8	1.6	2.4	0	0.8	1.6	2.4	0	0.8	1.6	2.4
Bicubic		-	30.39	29.42	27.24	25.37	27.55	26.84	25.42	24.09	27.21	26.72	25.52	24.41	24.46	24.02	22.95	21.89
RCAN [44]		167ms	**34.74**	32.90	29.12	26.75	**30.65**	29.49	26.75	24.99	**29.32**	28.56	26.55	25.18	**29.09**	26.89	23.85	22.30
SRMDNF [42] + Predictor [12]	×3	6ms	32.22	32.63	32.27	28.62	29.13	29.25	28.01	26.90	28.41	28.25	28.11	26.56	26.75	26.61	26.35	24.06
DASR (Ours)		70ms	34.11	**34.08**	**33.57**	**32.15**	30.13	**29.99**	**28.66**	**28.42**	28.96	**28.90**	**28.62**	**28.13**	27.65	**27.36**	**26.86**	**25.95**
Kernel Width			0	1.2	2.4	3.6	0	1.2	2.4	3.6	0	1.2	2.4	3.6	0	1.2	2.4	3.6
Bicubic		-	28.42	27.30	25.12	23.40	26.00	25.24	23.83	22.57	25.96	25.42	24.20	23.15	23.14	22.68	21.62	20.65
RCAN [44]		165ms	**32.63**	30.26	26.72	24.66	**28.87**	27.48	24.93	23.41	**27.72**	26.89	25.09	23.93	**26.61**	24.71	22.25	20.99
SRMDNF [42] + Predictor [12]	×4	5ms	30.61	29.35	29.27	28.65	27.74	26.15	26.20	26.17	27.15	26.15	26.15	26.14	25.06	24.11	24.10	24.08
IKC [12]		517ms	32.00	31.77	30.56	29.23	28.52	28.45	28.16	26.81	27.51	27.43	27.27	26.33	25.93	25.63	25.00	24.06
DASR (Ours)		70ms	31.99	**31.92**	**31.75**	**30.59**	28.50	**28.45**	**28.28**	**27.45**	27.51	**27.52**	**27.43**	**26.83**	25.82	**25.69**	**25.44**	**24.66**

veloped another variant (Model 3) by removing the channel-wise modulation coefficient branch. Note that, we adjust the number of channels in models 2 and 3 to ensure comparable model sizes. From Table 1 we can see that our DASR benefits from both dynamic convolutional kernels and channel-wise modulation coefficients to produce better results for various degradations.

Blind SR vs. Non-Blind SR. We further investigate the upper-bound performance of our DASR network by providing groundtruth degradation. Specifically, we replaced the degradation encoder with 5 FC layers to learn a representation directly from the true degradation (*i.e.*, blur kernel). This network variant (Model 5) was then trained from scratch for 500 epochs. When groundtruth degradation is provided, model 5 achieves improved performance and outperforms SRMDNF by notable margins. Further, SRMDNF is quite sensitive to degradation estimation errors under blind settings, with PSNR values being decreased if degradation is not accurately estimated (*e.g.*, 27.55 vs. 26.66/26.18 for $\sigma = 3.4$). In contrast, our DASR (Model 4) benefits from degradation representation learning to achieve better blind SR performance.

Study of Degradation Representations. Our degradation representations aim at extracting content-invariant degradation information from LR images. To demonstrate this, we conduct experiments to study the effect of different image contents to our degradation representations. Specifically, given an HR image, we first generated an LR image I_1 using a Gaussian kernel k. Then, we randomly selected another 9 HR images to generate LR images ($I_i(i = 2, 3, ...10)$) using k. Next, degradation representations were extracted from $I_i(i = 1, 2, ...10)$ to super-resolve I_1. Note that, $I_i(i = 2, 3, ...10)$ and I_1 share the same degradation but have different image contents. From Fig. 4 we can see that our network achieves relatively stable performance with
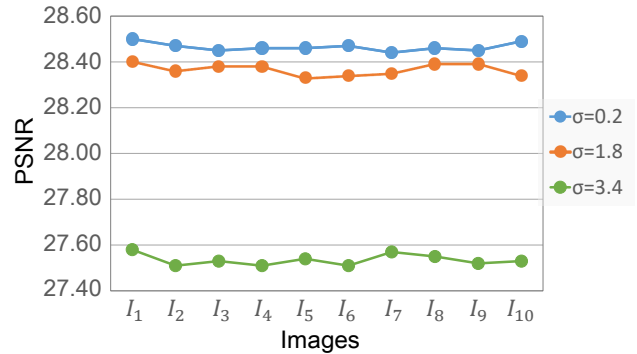

Figure 4. PSNR results achieved on Set14 using degradation representations learned from different image contents.

degradation representations learned from different image contents. This demonstrates that our degradation representations are robust to image content variations.

Comparison to Previous Networks. We compare DASR to RCAN, SRMD, MZSR and IKC. Pre-trained models of these networks are used for evaluation following their default settings. Quantitative results are shown in Table 2, while visualization results are provided in Fig. 5. Note that, MZSR[2]/IKC are only tested for ×2/4 SR since their pre-trained models for other scale factors are unavailable. For non-blind SR methods (SRMD and MZSR), we first performed degradation estimation to provide degradation information. Since KernelGAN is quite time-consuming (Table 1), the predictor sub-network in IKC was used to estimate degradations.

It can be observed from Table 2 that RCAN produces the highest PSNR results on bicubic degradation (*i.e.*, kernel width 0) while suffering relatively low performance when the test degradations are different from the bicubic one. Although SRMDNF and MZSR can adapt to the estimated

---

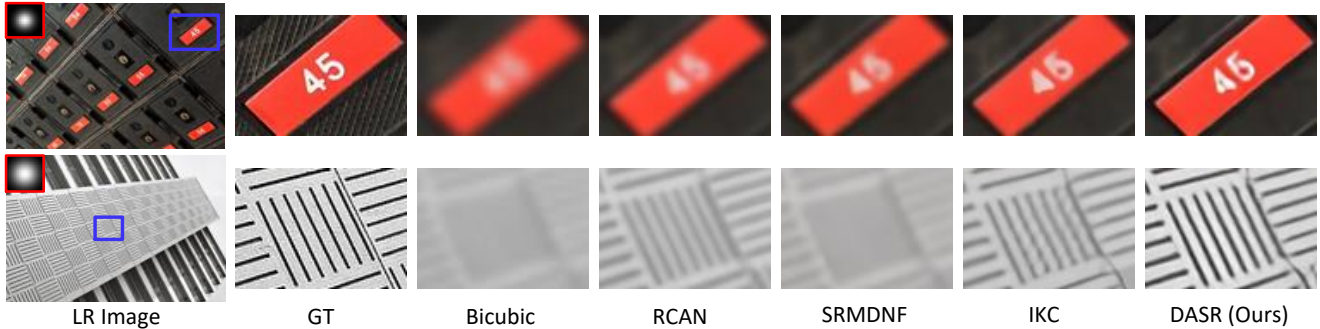[2]Pre-trained ×4 model of MZSR is based on s-fold downsampler.

Figure 5. Visual comparison achieved on Urban100 for ×4 SR. The blur kernels are illustrated with red boxes.

degradations, these methods are sensitive to degradation estimation, as demonstrated in Table 1. Therefore, degradation estimation errors can be magnified by SRMDNF and MZSR, resulting in limited SR performance. Since an iterative correction scheme is used to correct the estimated degradation, IKC outperforms SRMDNF with higher PSNR values being achieved. However, IKC is time-consuming due to its iterations. Compared to IKC, our DASR network achieves better performance for different degradations with shorter running time. That is because, our degradation representation learning scheme can extract "good" representations to distinguish different degradations at a single inference.

Visualization results achieved by different methods are shown in Fig. 5. Since RCAN is trained on the fixed bicubic degradation, it cannot reliably recover missing details when the real degradation differs from the bicubic one. Although SRMDNF can handle multiple degradations, failures can be caused by the degradation estimation error. By iteratively correcting the estimated degradations, IKC achieves better performance than SRMDNF. Compared to other methods, our DASR produces results with much clearer details and higher perceptual quality.

4.3. Experiments on General Degradations with Anisotropic Gaussian Kernels and Noises

We further conduct experiments on general degradations with anisotropic Gaussian kernels and noises. We first analyze the representations learned from general degradations and then compare the performance of our DASR to RCAN, SRMDNF and IKC under blind settings.

Study of Degradation Representations. Experiments are conducted to investigate the effect of two different components (*i.e.*, blur kernels and noises) to our degradation representations. We first visualize the representations for noise-free degradations with various blur kernels in Fig. 6(a). Then, we randomly select a blur kernel and visualize the representations for degradations with different noise levels in Fig. 6(b). It can be observed that our degradation encoder can easily cluster degradations with different noise levels into discriminative groups and roughly distinguish various
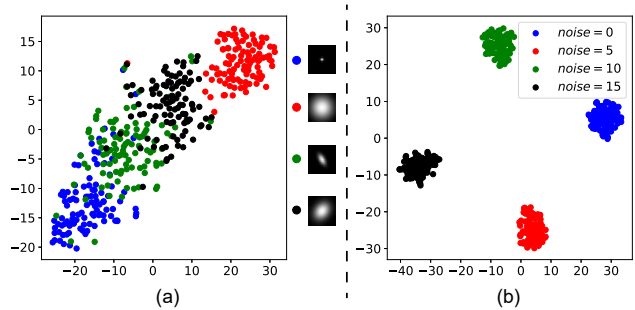

Figure 6. Visualization of representations for degradations with different blur kernels (a) and noise levels (b).

blur kernels.

Comparison to Previous Networks. We use 9 typical blur kernels and different noise levels for performance evaluation. To super-resolve noisy LR images using RCAN, SRMDNF and IKC, we first denoise the LR images using DnCNN [41] (a state-of-the-art denoising method) under blind settings. Since the pre-trained model of IKC is trained on isotropic Gaussian kernels only, we further fine-tuned this model on anisotropic Gaussian kernels for fair comparison. The predictor sub-network of the fine-tuned IKC model is used to estimate degradations for SRMDNF.

It can be observed from Table 3 that RCAN produces relatively low performance on complex degradations since it is trained on bicubic degradation only. Since SRMDNF is sensitive to degradation estimation errors, its performance for complex degradations is limited. By iteratively correcting the estimated degradations, IKC performs favorably against SRMDNF. However, IKC is more time-consuming since numerous iterations are required. Different from IKC that focuses on pixel-level degradation estimation, our DASR explores an effective yet efficient approach to learn discriminative representations to distinguish different degradations. Using our degradation representation learning scheme, DASR outperforms IKC in terms of PSNR for various blur kernels and noise levels with running time being reduced by over 7 times. Figure 7 further illustrates the visualization results produced by different methods. Our DASR achieves much better visual quality while

Table 3. PSNR results achieved on Set14 for ×4 SR.

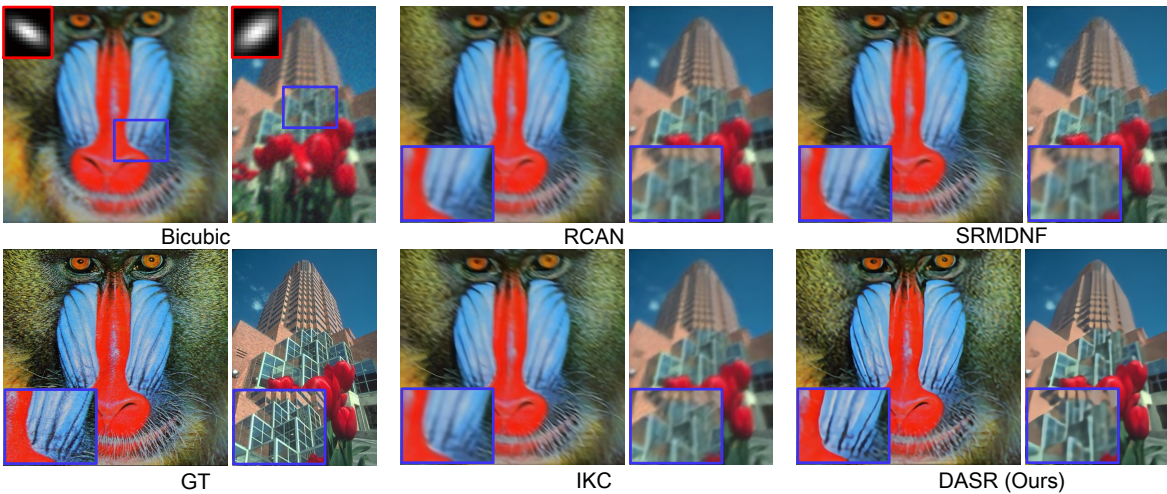| Method	#Params.	Time	Noise	Blur Kernel								
DnCNN [41]+RCAN [44]	650K+15.2M	169ms	0	26.44	26.22	24.48	24.23	24.29	24.19	23.90	23.42	23.01
			5	26.10	25.90	24.29	24.07	24.14	24.02	23.74	23.31	22.92
			10	25.65	25.47	24.05	23.84	23.92	23.80	23.54	23.14	22.77
DnCNN [41]+SRMDNF [44]+Predictor [12]	650K+1.5M+420K	8ms	0	26.84	26.88	25.57	25.69	25.64	24.98	25.12	25.28	24.84
			5	25.92	25.75	24.18	23.97	24.05	23.93	23.65	23.20	22.80
			10	25.39	25.23	23.88	23.68	23.74	23.65	23.39	22.99	22.64
DnCNN [41]+IKC [12]	650K+5.2M	520ms	0	27.71	27.78	27.11	27.02	26.93	26.65	26.50	26.01	25.33
			5	26.91	26.80	24.87	24.53	24.56	24.40	24.06	23.53	23.06
			10	26.16	26.09	24.55	24.33	24.35	24.17	23.92	23.43	23.01
DASR (Ours)	5.8M	70ms	0	**27.99**	**27.97**	**27.53**	**27.45**	**27.43**	**27.22**	**27.19**	**26.83**	**26.21**
			5	**27.25**	**27.18**	**26.37**	**26.16**	**26.09**	**25.96**	**25.85**	**25.52**	**25.04**
			10	**26.57**	**26.51**	**25.64**	**25.47**	**25.43**	**25.31**	**25.16**	**24.80**	**24.43**

Figure 7. Visual comparison achieved on Set14 and B100. Noise levels are set to 0 and 5 for these two images, respectively.

other methods suffer obvious blurring artifacts.

4.4. Experiments on Real Degradations

We further conduct experiments on real degradations to demonstrate the effectiveness of our DASR. Following [42], DASR trained on isotropic Gaussian kernels is used for evaluation on real images. Visualization results are shown in Fig. 8. It can be observed that our DASR produces visually more promising results with clearer details and fewer blurring artifacts.

5. Conclusion

In this paper, we proposed an unsupervised degradation representation learning scheme for blind SR to handle various degradations. Instead of explicitly estimating the degradations, we use contrastive learning to extract discriminative representations to distinguish different degradations. Moreover, we introduce a degradation-aware SR (DASR) network with flexible adaption to different degradations based on the learned representations. It is demonstrated that our degradation representation learning scheme can extract discriminative representations to obtain accurate
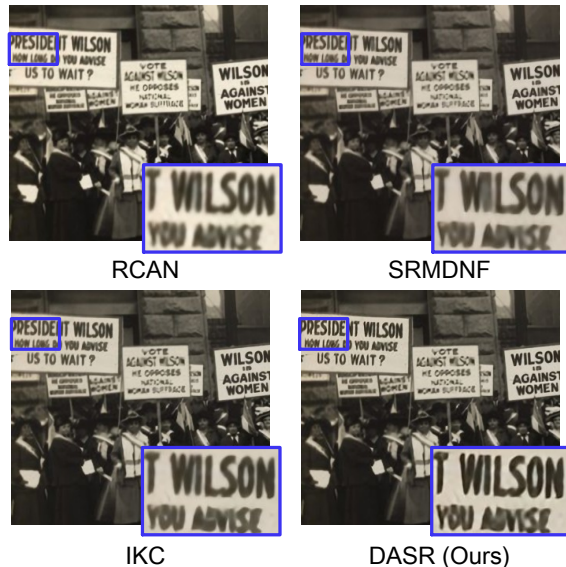

Figure 8. Visualization results achieved on a real image.

degradation information. Experimental results show that our network achieves state-of-the-art performance for blind SR with various degradations.

References

[1] Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, pages 1122–1131, 2017. 4

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *ECCV*, pages 252–268, 2018. 1

[3] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *NeurIPS*, pages 284–293, 2019. 1, 2, 3, 5

[4] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie-Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In *BMVC*, pages 1–10, 2012. 4

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020. 1, 2, 3

[6] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv*, 2020. 3

[7] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR*, 2019. 2

[8] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In *ICCV*, pages 1422–1430, 2015. 2

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, pages 184–199, 2014. 1, 2

[10] Alexey Dosovitskiy, Jost Tobias Springenberg, Martin Riedmiller, and Thomas Brox. Discriminative unsupervised feature learning with convolutional neural networks. In *NeurIPS*, pages 766–774, 2014. 1

[11] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. *ICLR*, 2018. 2

[12] Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *CVPR*, 2019. 1, 2, 3, 4, 5, 6, 8

[13] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *CVPR*, 2006. 1

[14] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *CVPR*, pages 1664–1673, 2018. 2

[15] Jingwen He, Chao Dong, and Yu Qiao. Modulating image restoration with continual levels via adaptive feature modification layers. In *CVPR*, pages 11056–11064, 2019. 4

[16] Jingwen He, Chao Dong, and Yu Qiao. Interactive multi-dimension modulation with dynamic controllable residual learning for image restoration. In *ECCV*, 2020. 4

[17] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross B. Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 2020. 1, 2, 3

[18] Olivier J Hénaff, Aravind Srinivas, Jeffrey De Fauw, Ali Razavi, Carl Doersch, SM Eslami, and Aaron van den Oord. Data-efficient image recognition with contrastive predictive coding. In *CVPR*, 2019. 2

[19] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *CVPR*, pages 5197–5206, 2015. 4

[20] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers. In *CVPR*, 2020. 2, 5

[21] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, pages 1646–1654, 2016. 2

[22] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *CVPR*, pages 1637–1645, 2016. 1

[23] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015. 5

[24] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPR*, 2017. 1, 2

[25] Zhengxiong Luo, Yang Huang, Shang Li, Liang Wang, and Tieniu Tan. Unfolding the alternating optimization for blind super resolution. In *NeurIPS*, 2020. 2, 5

[26] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(Nov):2579–2605, 2008. 5

[27] David Martin, Charless Fowlkes, Doron Tal, Jitendra Malik, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *ICCV*, 2001. 4

[28] Tomer Michaeli and Michal Irani. Nonparametric blind super-resolution. In *ICCV*, pages 945–952, 2013. 1, 2, 3

[29] Mehdi Noroozi, Hamed Pirsiavash, and Paolo Favaro. Representation learning by learning to count. In *ICCV*, pages 5898–5906, 2017. 2

[30] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv*, 2018. 2

[31] Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *ECCV*, 2020. 3

[32] Yajun Qiu, Ruxin Wang, Dapeng Tao, and Jun Cheng. Embedded block residual network: A recursive restoration model for single-image super-resolution. In *ICCV*, 2019. 1

[33] Assaf Shocher, Nadav Cohen, and Michal Irani. 'Zero-shot' super-resolution using deep internal learning. In *CVPR*, 2018. 1, 2, 5

[34] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In *CVPR*, 2020. 1, 2, 6

[35] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. *arXiv*, 2019. 1, 2

[36] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, Lei Zhang, and Bee Lim *et al.* NTIRE 2017 challenge on single image super-resolution: Methods and results. In *CVPRW*, pages 1110–1121, 2017. 4

[37] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *CVPR*, pages 3733–3742, 2018. 2

[38] Yu-Syuan Xu, Shou-Yao Roy Tseng, Yu Tseng, Hsien-Kai Kuo, and Yi-Min Tsai. Unified dynamic convolutional network for super-resolution with variational degradations. In *CVPR*, pages 12496–12505, 2020. 2, 4

[39] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *International Conference on Curves and Surfaces*, volume 6920, pages 711–730, 2010. 4

[40] Kai Zhang, Luc Van Gool, and Radu Timofte. Deep unfolding network for image super-resolution. In *CVPR*, 2020. 1, 2, 5

[41] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.*, 26(7):3142–3155, jul 2017. 7, 8

[42] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *CVPR*, 2017. 1, 2, 3, 4, 5, 6, 8

[43] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *ECCV*, pages 649–666, 2016. 2

[44] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, pages 1646–1654, 2018. 2, 3, 5, 6, 8

[45] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *CVPR*, pages 2472–2481, 2018. 2